中粒咖啡 *WRKY* 基因家族的生物信息学分析

付春[1,21]，唐雪[1,2]，杨瑶君[1,2]，江纳[3]

(1. 乐山师范学院 竹类病虫防控与资源开发四川省重点实验室，四川 乐山，614000；2. 乐山师范学院 生命科学学院，四川 乐山， 614000；3. 乐山师范学院 数学与信息科学学院，四川 乐山，614000)

摘要: *WRKY* 转录因子是植物信号网络中不可缺少的部分，作为植物中最大的转录因子家族之一，在植物的多种应激反应中发挥重要作用。利用生物信息学方法对中粒咖啡 WRKY 蛋白家族的理化特性及其分子进化进行了详细的分析。结果表明：CcWRKY 蛋白氨基酸数量在 103~994 个之间，均无信号肽，推测其为非分泌性蛋白。其二级结构以无规则卷曲为最主要的结构元件，三级结构主要分为 6 类，其中以 CcWRKY61、CcWRKY72、CcWRKY6 和 CcWRKY7 为主要成员的 D 类结构最稳定。保守结构域及进化树分析表明中粒咖啡 *WRKY* 基因家族含有 49 个成员，其中的 10 个成员归为 *WRKY* 第Ⅰ家族，34 个成员归为 *WRKY* 第Ⅱ家族，5 个成员归为 *WRKY* 第Ⅲ家族。中粒咖啡 *WRKY1* 基因与其他物种的系统进化分析结果表明其 *WRKY1* 与烟草亲缘关系最近，与非洲油棕(*Elaeis guineensis*)亲缘关系最远，说明 WRKY1 蛋白在生物进化过程中还是比较保守的。该文研究结果可以为中粒咖啡 *WRKY* 基因家族的分子功能的深层次研究提供一定的借鉴作用，对进一步探究中粒咖啡 *WRKY* 基因的功能、进化以及分子育种具有重要的现实意义。

关键词: 中粒咖啡，*WRKY*，转录因子，系统进化，生物信息学

中图分类号: Q943.2　　　　　　**文献标识码：A**

Bioinformatics analysis of *WRKY* gene family in *Coffee canephora*

FU Chun[1,2*], TANG Xue[1,2], YANG Yaojun[1,2], JIANG Na[3]

(1. *Bamboo Disease and Pests Control and Resources Development Key Laboratory of Sichuan Province*, *Leshan Normal University*, Leshan 614000, Sichuan, China; 2. *College of Life Science*, *Leshan Normal University*, Leshan 614000, Sichuan, China; 3. *College of Mathematics and Information Science*, *Leshan Normal University*, Leshan 614000, Sichuan, China)

Abstract: *WRKY* transcription factor is an indispensable part of plant signal network. As one of the largest family of transcription factors in plants, *WRKY* transcription factor plays an important role in various stress responses of plants. The physicochemical properties and molecular evolution of *WRKY* protein family of *Coffee canephora* were analyzed in detail by bioinformatics methods. The results showed that the amino acids numbers of WRKY proteins were from 103 to 994, and there was no signal peptide. It is presumed to be a non-secretory protein. Random coil is the most important structural element in the secondary structure of its WRKY proteins. The tertiary structure of WRKY proteins were

---

收稿日期: 2020-05-20

基金项目: 国家自然科学基金(31470655)；四川省科技厅项目(17ZZ019)；乐山市科技局重点研究项目(18JZD101)；乐山师范学院人才启动项目(XJR17005)；乐山师范学院重点研究项目(LZD010)；乐山师范学院横向项目(LHX190786)[Supported by the National Natural Science Foundation of China(31470655)；Sichuan Provincial Science and Technology Department Project (17ZZ019)；Key Research Project of Leshan Municipal Science and Technology Bureau (18JZD101)；Leshan Normal University Talent Launch Project (XJR17005)；Key Research Project of Leshan Normal University(LZD010),Cooperative Project of Leshan Normal University(LHX190786)]。

作者简介: 付春(1984-)，男，江西高安人，博士，讲师，从事经济植物的生物信息学研究，（E-mail）fuchun421@aliyun.com。

[*]通信作者

mainly divided into six categories. Among them, the D-class structure with CcWRKY61, CcWRKY72, CcWRKY6 and CcWRKY7 as the main members is the most stable. Conserved domain and phylogenetic tree analysis indicated that the *WRKY* gene family in *C. canephora* contains 49 members, 10 of which are classified as *WRKY* class I, 34 members are classified as *WRKY* class II, and 5 members are classified as *WRKY* class III. The phylogenetic analysis of *WRKY1* gene of *C. canephora* and other species showed that *WRKY1* of *C. canephora* was the most closely related to tobacco and the most distant to *Elaeis guineensis*, indicating that WRKY1 protein is relatively conservative in the process of biological evolution. The results of this study can provide some reference for the further study of the molecular function of *WRKY* gene family of *Coffea Canephoea*, and have important practical significance for further exploration of the function, evolution and molecular breeding of *WRKY* gene in *C. canephora*.

Key words: *Coffee canephora*, *WRKY*, transcription factor, phylogenetic evolution, bioinformatics

中粒咖啡(*Coffee canephora*)的年平均产量和产值均达到三大饮料作物之首，而且其消费量也居第一(闫林等，2018)。它作为一种重要的提神饮料已有成百上千年的历史，是国际贸易中非常重要的的原料产品(Davis et al., 2011; Monente et al., 2015)。中粒咖啡是属于茜草科咖啡属的一种小乔木或灌木。咖啡属植物共 100 多种，而生产性栽培的主要为小粒种和中粒种，分别占据世界咖啡产量的 60%和 40%。2014 年，海南以种植中粒种咖啡为主，咖啡种植面积占全国的 0.47% (Rodrigues et al., 2015; 闫林等，2018)。中粒种咖啡理想的种植环境是在气温高并且降雨量丰沛的地方，然而，在中粒咖啡的栽培过程中经常会遭遇病害、干旱和低温等胁迫，严重影响了生长发育，给生产带来重大损失(杨华庚，2014)。

在植物中，WRKY 蛋白是一种普遍存在的基因超家族转录因子，是植物科学研究中最多的一种转录因子，归属于高等植物的 10 大超家族转录因子(郑超等，2018)。*WRKY* 因其家族所有成员的 N 端均含有由 WRKYGQK 7 个氨基酸组成的保守序列而得名，依据其特点，将植物中的 *WRKY* 基因超家族(李智等，2019)划分为 3 个家族：I、II 和 III(郑超等，2018)。第 I 家族具有 2 个 WRKY 域和 1 个锌指结构为 C2H2(C-X4-5-C-X22-23-H-X1-H) (其中 X 可以表示任一氨基酸)，第 II 家族是具有 1 个 WRKY 域和 1 个以 C2H2(C-X4-5-C-X22-23-H-X1-H)为代表的锌指结构域组成的，第 III 家族是由 1 个 WRKY 域和 C2HC 型锌指结构(C-X7-C-X23-H-X1-C)组成(Bakshi et al., 2014)。Ishiguro & Nakamura(1994)首次在甘薯(*Ipomoea batatas*)克隆首个 *WRKY* 基因并鉴定了 1 个 DNA binding 蛋白，命名为 *SPF1* 转录因子。随后在不同的作物中克隆和鉴定到了大量的 *WRKY* 转录因子，包括：蓖麻(56 个 *WRKY* 基因)(邹智，2013)、猕猴桃(89 个 *WRKY* 基因)(包昌艳等，2018)、番茄(81 个 *WRKY* 基因)(张红，2017)、葡萄(56 个 *WRKY* 基因)(苏玲等，2019)和枣(92 个 *WRKY* 基因)(薛宝平等，2018)等。

WRKY 转录因子是一类植物特有的转录因子，它参与植物种子解除休眠、代谢和激素信号转导等调控(谢政文等，2016)。在很多植物的发育和生理过程中具有转录激活或转录抑制的功能(Rushton et al., 2010; Rushton et al., 2012; Ding et al., 2015)。有研究分析指出，*WRKY* 基因家族在盐胁迫、高温、低温和干旱等应答非生物胁迫条件下起到了非常重要的作用(Jiang et al., 2006; Wu et al., 2009; 李晓颖等，2019)。也有研究报道 *WRKY* 转录因子在植物的生长发育过程中发挥作用。比如果实成熟(Xu et al., 2004)、胚胎发育(Lagacé et al., 2004)和衰老(Robatzek et al., 2010; 张柳等，2019)等。在猕猴桃中有 33 个 *WRKY* 基因在植物根、叶、花和果 4 个器官中均有显著性表达(包昌艳等，2018)；在苹果 *WRKY* 基因家族研究中得到有 12 个 *MdWRKY* 均在根、茎、叶、花和果中表达，可能参与调控苹果生长和发育等过程(张柳等，2019)；在烟草‘云烟 87’叶片衰老中 *WRKY* 转录因子研究中指出，*NtWRKY11*、*NtWRKY31*、*NtWRKY40* 和 *NtWRKY51* 与逆境胁迫相关，又在叶片衰老中起到交叉作用(谷彦冰等，2015)；目前为止，在中粒咖啡转录因子中，有对 *TPS* 基因家族进行了系统分析(程甜等，2016)，而对 *WRKY* 基因家族研究甚少，Dong et al.(2019)对中粒咖啡全基因组鉴定了 49 个 *CcWRKYs*，并研究了其对寒冷胁迫的响应。本研究将通过生物信息学的方法对中粒咖啡 *WRKY* 转录因子的氨基酸的组成、跨膜结构域、保守基序、蛋白的高级结构分析

以及与烟草、猕猴桃、水稻等物种的亲缘关系进行预测和分析，为进一步探索与生物逆境相关的*WRKY* 转录因子的功能研究提供一定的理论依据。

1 材料与方法

1.1 试验材料

从植物转录因子数据库(http://planttfdb.cbi.pku.edu.cn/index.php)(数据库版本为 V5.0)获取中粒咖啡(*Coffea canephora*)、水稻(*Oryza sativa* subsp. *indica*)、玉米(*Zea mays*)、猕猴桃(*Actinidia chinensis*)、烟草(*Nicotiana tabacum*)、板栗(*Castanea mollissima*)、非洲油棕(*Elaeis guineensis*)、小果野蕉(*Musa acuminata*)、毛竹(*Phyllostachys heterocycla*)、枣(*Ziziphus jujuba*)、苹果(*Malus domestica*)和拟南芥(*Arabidopsis thaliana*)这 12 个物种的 *WRKY* 转录因子家族的核苷酸序列及氨基酸序列。通过 Blastp 程序比对拟南芥 WRKY 蛋白序列找出中粒咖啡转录因子 *WRKY(CcWRKY)*基因家族成员的对应关系(表 1)。

1.2 方法

1.2.1 *WRKY* 家族成员编码蛋白理化性质分析

运用 PrtoParam 在线程序分析了 CcWRKY 蛋白的理化性质，包括氨基酸长度、总平均亲疏水性及带正负电荷氨基酸数量等指标的相关预测。利用在线软件 SignalP-5.0 和 CELLO v2.5(Yu et al., 2006; 沙桐等，2019)分别对 49 条中粒咖啡 *WRKY* 基因家族的氨基酸序列进行信号肽预测和亚细胞定位分析。

1.2.2 氨基酸的跨膜结构、亲疏水性、磷酸化位点及保守基序预测

在 TMHMM Server v.2.0 在线软件(周洁等，2018)输入中粒咖啡 WRKY 蛋白序列，分析该蛋白家族的跨膜结构。通过 Protscale 在线网站完成其亲疏水性预测分析。在 NetPhos3.1 Server (http://www.cbs.dtu.dk/services/NetPhos/)对中粒咖啡 WRKY 蛋白序列潜在磷酸化位点进行预测分析。利用 MEME 5.0.4 在线工具搜索中粒咖啡 *WRKY* 基因家族成员的 Motif，参数设置：保守基序个数为 10，其他参数为默认设置。

1.2.3 蛋白质的二级结构和三级结构预测

利用 SOPMA 在线网站预测分析了中粒咖啡 49 种 WRKY 蛋白的二级结构(张春兰等，2018; 周洁等，2018)。利用 Phyre 2(Kelley et al., 2015)网站对 49 种 WRKY 蛋白的三级结构进行预测分析。

1.2.4 中粒咖啡 *WRKY* 基因家族系统发育树的构建

利用 MEGA 7.0 软件(王鹏洋等，2019)对 CcWRKY 蛋白序列进行多重比对并构建其系统发育树，并且对中粒咖啡、拟南芥、烟草、猕猴桃、板栗、苹果和玉米等 12 个物种的 WRKY1 蛋白序列进行系统发育树的构建，采用邻接法(Neighbor-Joining)构建系统发育树(Saitou et al., 1987; Kumar et al., 2018)，对生成的系统发育树进行校正，校验参数 Bootstrap 重复 1 000 次，得到最终的系统发育树。

2 结果与分析

2.1 中粒咖啡 WRKY 蛋白一级结构和理化特性

运用在线程序 PrtoParam(周洁等，2018；王鹏洋等，2019；沙桐等，2019)搜索出中粒咖啡转录因子 *WRKY* 基因家族中的共有 49 个成员，将这些成员的蛋白一级结构和理化性质进行汇总及分析。结果表示，中粒咖啡 *WRKY* 基因家族的蛋白的氨基酸总数在 103~994 个之间，氨基酸平均为 376 个。多数为酸性蛋白质，只有 19 个成员的理论等电点大于 8.0；不稳定系数显示 Cc02_g05280 和 Cc10_g06400 为稳定蛋白，其余均为不稳定蛋白(不稳定系数小于 40 的为稳定蛋白)。有 2 个成员的带负电荷的残基总数等于带正电荷的残基总数，分别是 Cc01_g14950 和 Cc07_g03120，显示电中性。有 25 个成员的带负电荷的残基总数大于带正电荷的残基总数，显示带负电荷；有 22 个成员带正电荷的残基总数大于带负电荷的残基总数，显示带正电荷(表 1)。

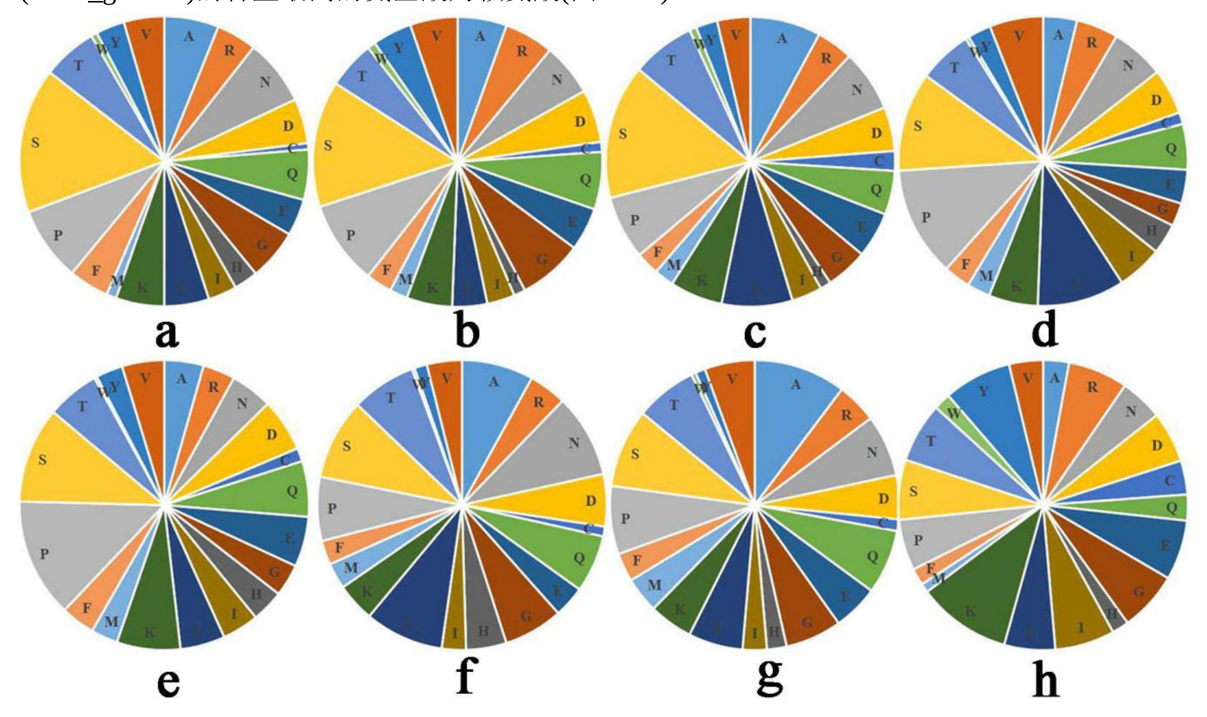
表 1 中粒咖啡 *WRKY* 基因家族的蛋白特征

Table 1 Protein characteristics of *WRKY* gene family in *Coffea canephora*

CcWRKY protein ID	CcWRKY 蛋白名称 CcWRKY protein name	氨基酸数目 Number of amino acid	等电点 pI	Asp+Glu	Arg+Lys	不稳定系数 Instability index	脂肪族氨基酸指数 Aliphatic index	亚细胞定位 Subcellular localization
Cc01_g14950	CcWRKY01	545	7.05	58	58	67.75	48.13	细胞核 Nuclear
Cc01_g18730	CcWRKY02	343	9.62	29	50	60.7	68.8	细胞核 Nuclear
Cc02_g01860	CcWRKY03	333	6.71	40	39	45.77	69.37	细胞核 Nuclear
Cc02_g02240	CcWRKY04	270	8.95	24	29	64.87	50.3	细胞核 Nuclear
Cc02_g05270	CcWRKY05	333	6.71	40	39	45.77	69.37	细胞核 Nuclear
Cc02_g05280	CcWRKY06	139	5.85	18	16	37.27	42.81	细胞核 Nuclear
Cc02_g15550	CcWRKY07	291	5.44	45	37	64.87	41.24	细胞核 Nuclear
Cc02_g22190	CcWRKY08	558	5.72	68	55	53.17	68.37	细胞核 Nuclear
Cc02_g25200	CcWRKY09	234	8.47	25	28	45.51	45.34	细胞核 Nuclear
Cc02_g38550	CcWRKY10	367	6.25	37	34	69.09	50	细胞核 Nuclear
Cc02_g38600	CcWRKY11	235	8.73	21	24	62.12	79.19	细胞核 Nuclear
Cc02_g39120	CcWRKY12	367	5.26	54	43	47.86	60.03	细胞核 Nuclear
Cc03_g00670	CcWRKY13	338	6.46	39	36	68.94	53.93	细胞核 Nuclear
Cc04_g05080	CcWRKY14	318	8.11	39	41	46.83	66.54	细胞核 Nuclear
Cc04_g15400	CcWRKY15	524	6.84	47	46	49.41	62.58	细胞核 Nuclear
Cc05_g04000	CcWRKY16	622	6.63	58	53	42.83	64.37	细胞核 Nuclear
Cc05_g08580	CcWRKY17	275	5.03	36	26	44.09	53.96	细胞核 Nuclear
Cc05_g09630	CcWRKY18	377	6.59	35	33	59.04	37.69	细胞核 Nuclear
Cc05_g14660	CcWRKY19	545	5.95	63	54	54.42	68.81	细胞核 Nuclear
Cc05_g15790	CcWRKY20	730	6.07	82	70	57.21	57.07	细胞核 Nuclear
Cc06_g02490	CcWRKY21	298	6.39	32	30	50.27	59.87	细胞核 Nuclear
Cc06_g03470	CcWRKY22	471	8.98	51	63	41.88	51.61	细胞核 Nuclear
Cc06_g15010	CcWRKY23	517	7.16	65	65	52.84	53.73	细胞核 Nuclear
Cc07_g00980	CcWRKY24	199	9.6	18	28	47.71	62.21	细胞核 Nuclear
Cc07_g03120	CcWRKY25	507	6.99	51	51	55.6	62.6	细胞核 Nuclear
Cc07_g03730	CcWRKY26	994	8.7	103	114	52.98	67.42	细胞核 Nuclear
Cc07_g06340	CcWRKY27	363	9.23	35	47	43.06	72.23	细胞核 Nuclear
Cc07_g16400	CcWRKY28	103	9.54	10	18	49.64	41.46	细胞核 Nuclear
Cc07_g16570	CcWRKY29	217	9.04	25	31	50.11	55.71	细胞核 Nuclear
Cc08_g06420	CcWRKY30	336	9.57	28	43	43.23	59.23	细胞核 Nuclear
Cc08_g07850	CcWRKY31	286	5.68	45	36	42.87	71.29	细胞核 Nuclear
Cc08_g07860	CcWRKY32	245	8.23	29	31	43.54	70.82	细胞核 Nuclear
Cc08_g11060	CcWRKY33	183	5.47	26	22	51.7	35.63	细胞核 Nuclear
Cc08_g15270	CcWRKY34	352	5.73	46	38	55.81	61.22	细胞核 Nuclear
Cc08_g15290	CcWRKY35	189	9.25	16	25	44.17	67.62	细胞核 Nuclear
Cc08_g15910	CcWRKY36	342	5.74	37	30	57.86	62.19	细胞核 Nuclear
Cc09_g01430	CcWRKY37	554	6.48	55	52	46.06	61.01	细胞核 Nuclear
Cc09_g03640	CcWRKY38	485	6.2	62	55	54.42	58，08	细胞核 Nuclear
Cc09_g07330	CcWRKY39	454	5.68	57	50	52.18	61.26	细胞核 Nuclear

Cc09_g08300	CcWRKY40	180	8.43	15	17	73.83	51.39	细胞核 Nuclear
Cc09_g09190	CcWRKY41	304	6.56	37	35	66.02	55.16	细胞核 Nuclear
Cc10_g04810	CcWRKY42	362	9.54	31	46	66.64	59.06	细胞核 Nuclear
Cc10_g06400	CcWRKY43	105	9.05	13	18	37.63	62.19	细胞核 Nuclear
Cc11_g12210	CcWRKY44	280	6.8	32	31	50.57	48.39	细胞核 Nuclear
Cc11_g12480	CcWRKY45	526	7.7	54	55	61.09	53.78	细胞核 Nuclear
Cc11_g17170	CcWRKY46	191	9.54	16	25	41.22	56.13	细胞核 Nuclear
Cc00_g06830	CcWRKY47	561	6.79	72	69	46.12	61.28	细胞核 Nuclear
Cc00_g13890	CcWRKY48	557	8.61	51	55	55.62	51.33	细胞核 Nuclear
Cc00_g21560	CcWRKY49	534	7.52	51	52	52.28	59.06	细胞核 Nuclear

根据这些成员的氨基酸组成进行分析比较，结果显示，其中有 44 个成员的含量最高的氨基酸为丝氨酸，含量在 7.5%到 16.6%之间(图 1：a，b，c)。其中有 2 个成员(Cc02_g38600 和 Cc03_g00670)的含量最高的氨基酸为脯氨酸(图 1：d，e)，1 个成员(Cc05_g04000)的含量最高的氨基酸为天冬酰胺(图 1：f)，1 个成员(Cc09_g01430)的含量最高的氨基酸为丙氨酸(图 1：g)，1 个成员(Cc10_g06400)的含量最高的氨基酸为赖氨酸(图 1：h)。



A. 丙氨酸; **R**. 精氨酸; **N**. 天冬酰胺; **D**. 天冬氨酸; **C**. 半胱氨酸; **Q**. 谷氨酰胺; **E**. 谷氨酸; **G**. 甘氨酸; **H**. 组氨酸; **I**. 异亮氨酸; **L**. 亮氨酸; **K**. 赖氨酸; **M**. 蛋氨酸; **F**. 苯丙氨酸; **P**. 脯氨酸; **S**. 丝氨酸; **T**. 苏氨酸; **W.** 色氨酸; **Y**. 酪氨酸; **V**. 缬氨酸; a,b,c,d,e,f,g,h 分别代表 Cc00_g13890, Cc01_g14950, Cc07_g03120, Cc02_g38600, Cc03_g00670, Cc05_g04000, Cc09_g01430 和 Cc10_g06400。

A. Alanine; **R**. Arginine; **N**. Asparagine; **D**. Aspartic acid; **C**. Cysteine; **Q**. Glutarnine; **E**. Glutamicacid; **G**. Glycine; **H**. Histidine; **I.** Isoleucine; **L**. Leucine; **K**. Lysine; **M**. Methionine; **F**. Phenylalanine; **P**. Proline; **S**. Serine;**T**. Threonine; **W**. Tryptophan; **Y**. Tyrosine; a, b, c, d, e, f, g, h represent Cc00_g13890, Cc01_g14950, Cc07_g03120, Cc02_g38600, Cc03_g00670, Cc05_g04000, Cc09_g01430 and Cc10_g06400, respectively.

图 1 氨基酸组成

Fig. 1 Amino acid composition

2.2 氨基酸的亲疏水性预测

借助于在线工具 Protscale 对该蛋白进行亲疏水性预测分析，总体分布来看，这些转录因子氨基酸的分值均为负值，表明它们都是亲水性蛋白。可以看出 49 种 WRKY 的平均亲水系数介于 -1.196 和-0.473 之间，表明其都是亲水性蛋白。最大疏水性数值范围是 0.289 和 2.589 之间，最

大的是 Cc07_g03730 成员，最小的是 Cc08_g11060 成员；最小疏水性数值范围是-3.967 和-2.511 之间，最大的是 Cc02_g05280 成员，最小的是 Cc06_g03470 成员。其中 Cc08_g11060 的平均亲水系数最小，说明其亲水性最强。它的最大疏水性位置是第 170 个，值为 0.589；最小疏水性位置是 104 个，值为-3.811(表 2)。

表 2 中粒咖啡 *WRKY* 转录因子亲疏水性预测

Table 2 Prediction of hydrophilicity and hydrophobicity of *WRKY* transcription factors in *Coffea canephora*

蛋白名称 Protein name	总平均疏水指数 Total average hydrophobic index	最大疏水性 Maximum hydrophobicity		最小疏水性 Minimum hydrophobicity	
		位置 Position	数值 Value	位置 Position	数值 Value
Cc00_g06830	-0.899	25	2.544	424	-2.989
Cc00_g13890	-0.848	87	1.856	350	-3.400
Cc00_g21560	-0.678	435	2.278	94	-3.256
Cc01_g18730	-0.929	18, 19	1.689	28, 29	-3.200
Cc01_g14950	-0.759	86	1.889	121	-3.344
Cc02_g01860	-0.579	341	1.344	116	-2.833
Cc02_g02240	-0.847	150	1.700	166, 167	-3.778
Cc02_g05270	-0.579	30	1.578	100	-2.856
Cc02_g05280	-0.856	106	1.289	17	-2.511
Cc02_g15550	-1.030	204	1.456	277	-3.911
Cc02_g22190	-0.618	469	2.311	30, 31	-3.544
Cc02_g25200	-1.036	46	1.378	168	-2.622
Cc02_g38550	-0.821	98	1.778	166, 167	-3.589
Cc02_g38600	-0.526	72	1.333	161	-2.689
Cc02_g39120	-0.674	58	1.467	108	-2.967
Cc03_g00670	-0.875	62	1.711	158	-3.833
Cc04_g05080	-0.810	301	1.800	96	-3.167
Cc04_g15400	-0.628	429	2.278	144	-2.800
Cc05_g04000	-0.738	570	2.367	234, 235	-3.511
Cc05_g08580	-0.624	235	1.856	46	-2.900
Cc05_g09630	-0.958	268	1.411	164	-3.444
Cc05_g14660	-0.592	460	1.744	254	-2.800
Cc05_g15790	-0.765	150	1.356	468	-3.311
Cc06_g02490	-0.804	29	1.333	196, 270	-2.967
Cc06_g03470	-0.908	31	1.533	356	-3.967
Cc06_g15010	-0.913	64, 65	1.656	338	-3.444
Cc07_g00980	-0.854	162	1.411	102	-3.333
Cc07_g03120	-0.644	419, 420	2.122	72	-3.078
Cc07_g03730	-0.523	246	2.789	847	-3.011
Cc07_g06340	-0.473	278	1.578	49	-2.856
Cc07_g16400	-0.925	89	1.011	54	-3.067
Cc07_g16570	-0.858	44	1.700	203	-2.689
Cc08_g06420	-0.601	5	1.622	202	-3.144

Cc08_g07850	-0.776	269	2.156	231	-2.844
Cc08_g07860	-0.621	514	2.156	345, 346	-2.844
Cc08_g11060	-1.196	170	0.589	104	-3.811
Cc08_g15270	-0.691	307	1.767	144	-3.967
Cc08_g15290	-0.614	53	1.456	118	-2.622
Cc08_g15910	-0.709	252	2.267	219, 220	-3.222
Cc09_g01430	-0.614	524	2.344	184, 185	-3.189
Cc09_g03640	-0.927	211	1.178	452	-3.489
Cc09_g07330	-0.782	201	1.578	412	-3.311
Cc09_g08300	-0.678	120	1.289	106	-2.978
Cc09_g09190	-0.826	243	1.256	13	-3.556
Cc10_g04810	-0.742	278	1.544	115, 116	-2.978
Cc10_g06400	-0.909	95	1.089	42	-2.744
Cc11_g12210	-0.802	11	0.956	119	-2.900
Cc11_g12480	-0.832	116	1.433	378	-3.256
Cc11_g17170	-0.732	55	1.722	119	-2.689

2.3 信号肽预测及亚细胞定位结果

这 49 个 WRKY 蛋白均无信号肽，推测其为非分泌性蛋白。分别对 49 种 WRKY 蛋白进行了亚细胞定位预测，这些家族成员的亚细胞定位，都位于细胞核(表 1)，说明 *CcWRKY* 家族成员蛋白亚细胞定位于细胞核。

2.4 编码氨基酸的跨膜结构分析

采用 TMHMM Server(v.2.0)在线程序(郑超等，2018)分析中粒咖啡 WRKY 蛋白的跨膜结构。可以看到 WRKY 中只有 1 个成员 Cc02_g22190 蛋白质序列的第 536 个到第 553 个氨基酸之间有 1 个明显跨膜区域(图 2)，其他的 48 个成员都没有跨膜区域，推断中粒咖啡 WRKY 不属于跨膜蛋白。

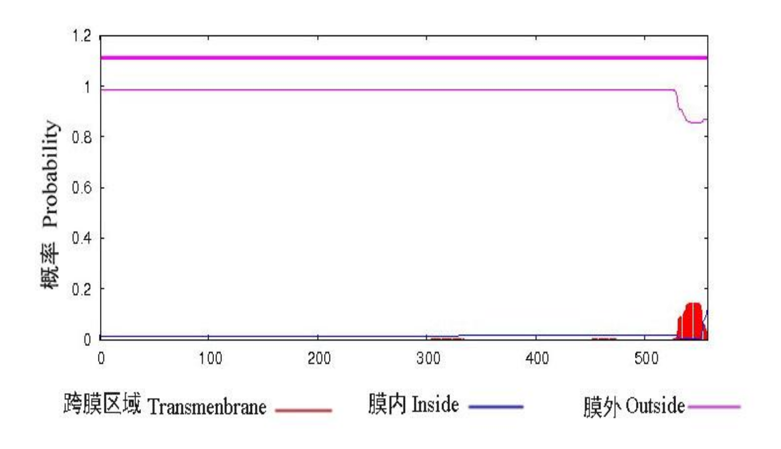


图 2 中粒咖啡 WRKY(Cc02_g22190)跨膜结构预测

Fig. 2 Prediction of the transmembrane structure of WRKY(Cc02_g22190) in *Coffea canephora*

2.5 磷酸化位点分析

经分析发现，这 49 种 WRKY 蛋白中丝氨酸磷酸化位点数量最多，其次是苏氨酸和酪氨酸。最有可能是潜在磷酸化位点的氨基酸是丝氨酸，其最小值已经达到 0.849，远远超过标准值 0.5，同时这一结果也说明 *CcWRKY* 基因家族蛋白可能通过相应位点的磷酸化来实现其功能的调控。其

中 Cc07_g03730 成员的丝氨酸、苏氨酸和酪氨酸磷酸化位点数量是最多的，分别有 83 个、39 个和 12 个(表 3)。

表 3 中粒咖啡 49 种 WRKY 蛋白磷酸化位点预测

Table 3 Prediction of phosphorylation sites of 49 WRKY proteins in *Coffea canephora*

蛋白名称 Protein name	磷酸化位点数量(个) Number of phosphorylation sites (number)			丝氨酸 Serine (S)		苏氨酸 Threonine (T)		酪氨酸 Tyrosine (Y)	
	丝氨酸 Serine (S)	苏氨酸 Threonine (T)	酪氨酸 Tyrosine (Y)	最大值 Max.	位置 Position	最大值 Max.	位置 Position	最大值 Max.	位置 Position
Cc00_g06830	44	20	6	0.998	6	0.996	387	0.834	240
Cc00_g13890	69	21	8	0.998	368	0.968	374	0.953	229
Cc00_g21560	63	19	3	0.998	152	0.909	176	0.922	390
Cc01_g14950	57	19	12	0.972	185	0.959	367	0.984	528
Cc01_g18730	31	6	1	0.981	223	0.967	69	0.742	284
Cc02_g01860	41	17	4	0.995	80	0.973	175	0.786	313
Cc02_g02240	33	10	5	0.993	77	0.920	254	0.985	89
Cc02_g05270	30	13	1	0.997	94, 40	0.963	157	0.558	180
Cc02_g05280	12	6	2	0.967	125	0.920	61	0.977	12
Cc02_g15550	27	15	1	0.996	20, 43	0.954	138	0.764	210
Cc02_g22190	35	26	3	0.996	65	0.982	47	0.755	194
Cc02_g25200	23	9	2	0.997	218	0.867	209	0.762	177
Cc02_g38550	43	13	2	0.995	152, 266	0.981	296	0.742	193
Cc02_g38600	21	6	2	0.997	116	0.811	138	0.997	166
Cc02_g39120	36	15	4	0.998	176	0.958	160	0.955	301
Cc03_g00670	24	15	4	0.996	218	0.975	239	0.920	308
Cc04_g05080	32	11	3	0.997	101	0.836	227	0.838	164
Cc04_g15400	62	26	4	0.992	63	0.991	46	0.912	106
Cc05_g04000	38	22	0	0.998	281	0.906	329	——	——
Cc05_g08580	23	19	3	0.994	29, 40	0.917	135	0.980	181
Cc05_g09630	34	13	8	0.993	166	0.811	30	0.929	304
Cc05_g14660	50	21	4	0.995	177	0.982	260	0.953	198
Cc05_g15790	57	23	6	0.994	355	0.932	518	0.942	319
Cc06_g02490	36	10	4	0.991	280	0.920	177	0.940	118
Cc06_g03470	35	19	6	0.992	258	0.906	55	0.823	439
Cc06_g15010	47	16	4	0.995	18	0.829	203	0.974	48
Cc07_g00980	14	7	2	0.994	159	0.870	91	0.820	177
Cc07_g03120	55	20	4	0.997	376	0.909	170	0.781	287
Cc07_g03730	83	39	12	0.997	138, 988	0.988	872	0.931	576
Cc07_g06340	39	13	3	0.990	45	0.908	12	0.816	292
Cc07_g16400	9	3	2	0.954	11	0.652	66	0.987	62
Cc07_g16570	14	8	3	0.991	162	0.951	54	0.762	144
Cc08_g06420	42	11	3	0.995	227, 207	0.941	231	0.875	331
Cc08_g07850	25	8	2	0.997	129	0.798	16	0.950	114

Cc08_g07860	21	4	1	0.996	31	0.570	101	0.814	91
Cc08_g11060	18	5	6	0.995	93	0.900	113	0.988	166
Cc08_g15270	33	16	4	0.992	278, 341	0.979	45	0.846	23
Cc08_g15290	10	8	2	0.993	105	0.863	106	0.762	114
Cc08_g15910	27	14	10	0.996	174	0.860	184	0.947	10
Cc09_g01430	34	15	0	0.996	280	0.936	273	——	——
Cc09_g03640	36	14	5	0.998	171	0.871	267	0.963	61
Cc09_g07330	47	21	1	0.997	446	0.935	278	0.546	56
Cc09_g08300	17	8	2	0.972	157	0.975	71	0.920	150
Cc09_g09190	24	14	7	0.969	204	0.978	192	0.829	10
Cc10_g04810	40	9	3	0.951	181	0.676	129	0.839	52
Cc10_g06400	5	5	1	0.849	37	0.741	16	0.973	67
Cc11_g12210	27	12	1	0.879	135	0.893	133	0.663	144
Cc11_g12480	55	20	2	0.969	214	0.894	169	0.845	237
Cc11_g17170	17	9	2	0.996	152	0.822	80	0.865	100

2.6 编码蛋白的二、三级结构预测分析

蛋白质的二级结构主要是指蛋白质的多肽链中有规则重复的构象，从 *CcWRKY* 家族成员的蛋白二级结构分析可以看出，*CcWRKY* 基因家族的二级结构主要由无规则卷曲、α-螺旋、β-转角和延伸链组成，这 4 种蛋白质的二级结构元件通常被主要研究(张春兰等，2018)，通过预测和数据分析将更容易了解到蛋白的空间结构。其中无规卷曲所占比例较高，其功能主要起连接其他二级结构元件，达到了 46.67%~77.26%，β-转角在二级结构中占比均较小。*CcWRKY* 基因家族的二级结构较整齐，除 Cc00_g06830、Cc05_g14660、Cc07_g16400、Cc08_g11060 和 Cc10_g06400 是无规则卷曲>延伸链>α-螺旋>β-转角外，其余基因家族成员均是无规则卷曲>α-螺旋>延伸链>β-转角(表 4，图 3)。

表 4 中粒咖啡 *WRKY* 基因家族的二级结构分析统计表

Table 4 Statistical table of secondary structure analysis of *WRKY* gene family in *Coffea canephora*

蛋白名称 Protein name	α-螺旋(X%)^{A} α-helix	延伸链(X%)^{B} Extended chain	β-转角(X%)^{C} β-turn	无规则卷曲(X%)^{D} Random coil
Cc00_g06830	63(11.23%)	79(14.08%)	31(5.53%)	388(69.16%)
Cc00_g13890	67(12.03%)	57(10.23%)	17(3.05%)	416(74.69%)
Cc00_g21560	110(20.60%)	61(11.42%)	12(2.25%)	351(65.73%)
Cc01_g14950	58(10.64%)	53(9.72%)	14(2.57%)	420(77.06%)
Cc01_g18730	86(25.07%)	31(9.04%)	12(3.50%)	214(62.39%)
Cc02_g01860	76(20.16%)	41(10.88%)	14(3.71%)	246(65.25%)
Cc02_g02240	69(25.56%)	13(4.81%)	9(3.33%)	179(66.30%)
Cc02_g05270	89(26.73%)	30(9.01%)	12(3.60%)	202(60.66%)
Cc02_g05280	31(22.30%)	19(13.67%)	13(9.35%)	76(54.68%)
Cc02_g15550	42(14.43%)	25(8.59%)	8(2.75%)	216(74.23%)
Cc02_g22190	174(31.18%)	59(10.57%)	12(2.15%)	313(56.09%)
Cc02_g25200	49(20.94%)	41(17.52%)	13(5.56%)	131(55.98%)
Cc02_g38550	100(27.25%)	30(8.17%)	14(3.81%)	223(60.76%)
Cc02_g38600	67(27.23%)	46(19.57%)	11(4.68%)	114(48.51%)
Cc02_g39120	71(19.35%)	40(10.90%)	12(3.27%)	244(66.49%)

Cc03_g00670	58(17.16%)	41(12.13%)	19(5.62%)	220(65.09%)
Cc04_g05080	93(29.25%)	38(11.95%)	4(1.26%)	183(57.55%)
Cc04_g15400	97(18.51%)	62(11.83%)	10(1.91%)	355(67.75%)
Cc05_g04000	147(23.63%)	74(11.90%)	14(2.25%)	387(62.22%)
Cc05_g08580	29(10.55%)	28(10.18%)	8(2.91%)	210(76.36%)
Cc05_g09630	60(17.80%)	33(9.79%)	14(4.15%)	230(68.25%)
Cc05_g14660	33(6.06%)	75(13.76%)	19(3.49%)	418(76.70%)
Cc05_g15790	81(11.10%)	65(8.90%)	20(2.74%)	564(77.26%)
Cc06_g02490	72(24.16%)	36(12.08%)	11(3.69%)	179(60.07%)
Cc06_g03470	74(15.71%)	50(10.62%)	20(4.25%)	327(69.43%)
Cc06_g15010	100(19.34%)	63(12.19%)	22(4.26%)	332(64.22%)
Cc07_g00980	60(30.15%)	30(15.08%)	18(9.05%)	91(45.73%)
Cc07_g03120	108(21.30%)	52(10.26%)	14(2.76%)	333(65.68%)
Cc07_g03730	267(26.86%)	178(17.91%)	78(7.85%)	471(47.38%)
Cc07_g06340	81(22.31%)	24(6.61%)	9(2.48%)	249(68.60%)
Cc07_g16400	13(12.62%)	23(22.33%)	12(11.65%)	55(50.40%)
Cc07_g16570	59(27.19%)	28(12.90%)	8(3.69%)	122(56.22%)
Cc08_g06420	71(21.13%)	35(10.42%)	16(4.76%)	214(63.69%)
Cc08_g07850	92(32.17%)	34(11.89%)	10(3.50%)	150(52.45%)
Cc08_g07860	61(24.90%)	30(12.24%)	6(2.5%)	148(60.41%)
Cc08_g11060	27(14.75%)	38(20.77%)	17(9.29%)	101(55.19%)
Cc08_g15270	100(28.41%)	30(8.52%)	15(4.26%)	207(58.81%)
Cc08_g15290	41(21.69%)	24(12.70%)	10(5.29%)	114(60.32%)
Cc08_g15910	72(21.05%)	37(10.82%)	11(3.22%)	222(64.91%)
Cc09_g01430	154(27.80%)	51(9.21%)	16(2.89%)	333(60.11%)
Cc09_g03640	65(13.40%)	58(11.96%)	16(3.30%)	346(71.34%)
Cc09_g07330	93(20.48%)	44(9.6%)	17(3.74%)	300(66.08%)
Cc09_g08300	33(18.33%)	20(11.11%)	9(5.00%)	118(65.56%)
Cc09_g09190	64(21.05%)	26(8.55%)	4(1.32%)	210(69.08%)
Cc10_g04810	75(20.72%)	34(9.39%)	10(2.76%)	243(67.13%)
Cc10_g06400	18(17.14%)	25(23.81%)	13(12.38%)	49(46.67%)
Cc11_g12210	38(13.57%)	31(11.07%)	12(4.29%)	199(71.07%)
Cc11_g12480	73(13.88%)	60(11.41%)	13(2.47%)	380(72.24%)
Cc11_g17170	46(24.08%)	31(16.23%)	11(5.76%)	103(53.93%)

注：[A] 表示 α-螺旋占该蛋白总二级结构的比例; [B] 表示延伸链占该蛋白总二级结构的比例; [C] 表示 β-转角占该蛋白总二级结构的比例; [D] 表示无规则卷曲占该蛋白总二级结构的比例。

Note: [A] represents the ratio of α-helix to the total secondary structure of the protein, [B] represents the ratio of the extended chain to the total secondary structure of the protein, [C] represents the ratio of the β-turn to the total secondary structure of the protein, and [D] represents the random coil to the total secondary structure of the protein.

分析发现中粒咖啡 *WRKY* 转录因子基因家族蛋白的三级结构可以分为 6 类，分别以 A、B、C、D、E 和 F 表示。其中，D 类成员最多，有 25 个；其次是 A 类，有 9 个成员；E 类含有 7 个成员，B 类含有 4 个成员，C 类含有 3 个成员；最少的是 F 类，只有一个成员。从三级结构稳定性来看，结构稳定从强到弱的顺序为 F>D、E>A>C>B。其中，Cc07_g03730 以 c5yvfD 结构为模板，有 315 个氨基酸残基与模板高达 100%的可信度；Cc05_g08580 以 c2aydA 结构为模板，有 75

个氨基酸残基与模板达到100%的可信度(图 4)。

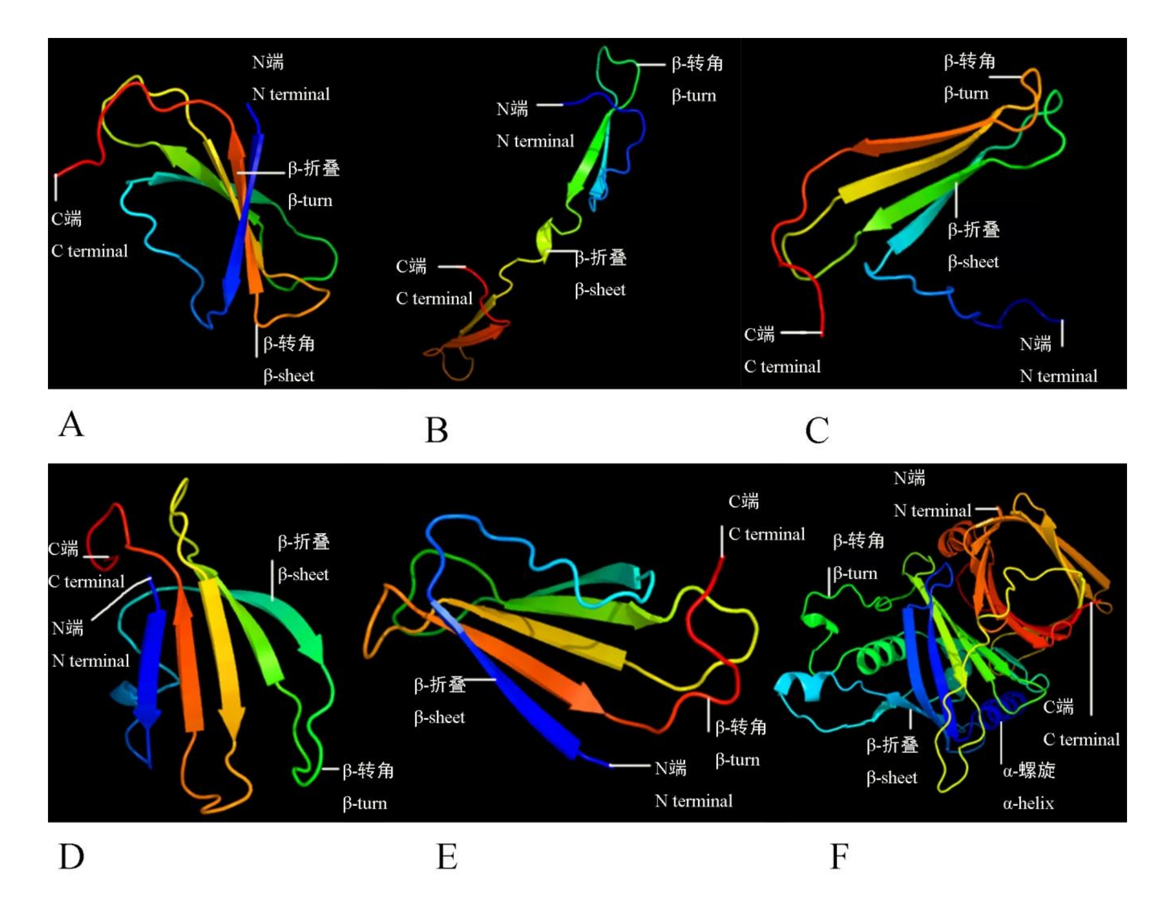


红色线条代表延伸链；蓝色线条代表 α-螺旋；紫色线条代表无规则卷曲；绿色线条代表 β-转角。

Red lines represent extended strand; Blue lines represent α-helix; Purple lines represent random coil; Green lines represent β-turn.

图 3 中粒咖啡 *WRKY* 基因家族二级结构分析

Fig. 3 Secondary structure analysis of *WRKY* gene family in *Coffea canephora*



A, **B**, **C**, **D**, **E**, **F** 分别表示 Cc05_g08580, Cc02_g05270, Cc09_g07330, Cc06_g02490, Cc03_g00670, Cc07_g03730.

A, **B**, **C**, **D**, **E**, **F** represent Cc05_g08580, Cc02_g05270, Cc09_g07330, Cc06_g02490, Cc03_g00670, Cc07_g03730, respectively.

图 4 三级结构模型

Fig. 4 Tertiary structure model

2.7 氨基酸序列结构域进行分析

借助于 MEME 在线工具搜索中粒咖啡 *WRKY* 基因家族成员的 Motif，根据其保守性的强弱，依次命名为 Motif1~Motif6。由序列 logo 可以看出，WRKYGQK 七肽序列的基序分别存在于 Motif 1 和 Motif 5 中，*WRKY* 基因家族 49 个成员均具有 Motif1 和 Motif2，推测 Motif1 是 WRKY 结构域。每个成员至少都含有 1 个 WRKY 七肽，其中最多的有 5 个保守基序，最少的有 2 个保守基序。有 19 个成员只有 Motif1 和 Motif2；有 7 个成员含有 Motif1、Motif2、Motif4、Motif5 和 Motif6(图 5，表 5)。

2.8 中粒咖啡 WRKY 蛋白的系统发育树的构建

对 *WRKY* 基因家族成员蛋白序列进行多重比对分析，将下载的 49 个家族成员蛋白序列输入 MEGA 绘制出中粒咖啡 WRKY 蛋白的进化树(图 6)。从进化树的聚类结果可以清晰地看出这 49 个成员分成了 6 组(依次称为 Group 1, Group 2, Group 3, Group 4, Group 5 和 Group 6)。其中 Group 1 成员最多[14 个, WRKY 12, 13, 23, 28, 48, 50, 51 (Cc07-g16400 和 Cc10-g06400), 56 (Cc02-g38600 和 Cc08-g15290), 57, 71 和 75 (Cc07-g00980 和 Cc11-g17170)]，约占总数的 28.4%；其次是 Group 2 [10 个, WRKY 1, 2, 4, 20, 32, 33 (Cc00-g13890 和 Cc01-g14950), 34 和 44]、Group 5 [10 个, WRKY 7 (Cc07-g0634 和 Cc10-g04810), 11, 21, 22 (Cc02-g38500 和 Cc08-g15270), 27, 35, 65 和 69]和 Group 3 [9 个, WRKY 7 (Cc05-g04000 和 Cc09-g01430), 9, 40 (Cc04-g05080, Cc08-g07850 和 Cc08-g07860), 61 和 72(Cc00-g21560 和 Cc07-g03120)]；剩余 2 组成员均较少，Group 6 有 5 个[WRKY 41 (Cc02-g01820, Cc02-g39120 和 Cc08-g15910), 55 和 70]，而 Group 4 仅有 1 个(WRKY 49) (表 1)。从进化树分支关系中还可以得出 Group 1、Group 2、Group 3、Group 4 和 Group 5 聚类到为一大组，说明它们的亲缘关系较近，而 Group 6 与它们的关系较远，因此，推测 Group 6 成员可能是

在物种长期遗传进化过程中非常保守，且与祖先关系最相似的一类而被保留下来的。

分析进化树分组结果后，采取家族成员分组比对的方法，发现每组的保守氨基酸呈多样性，但值得惊奇的是所有组都具有 WRKY 保守基序，参考 Eulgem et al.(2000)分类标准，将得到的 49 个 *WRKY* 基因家族成员分为 3 类：Group 2 (D731-H785、D943-H998)属于 WRKY 第Ⅰ家族，它含有 2 个保守的七肽结构(WRKYGQK)和 2 个锌指结构(C2H2 型)；Group 6 (D151-C209)属于 WRKY 第Ⅲ家族，它含有 1 个保守的七肽结构(WRKYGQK)和 1 个锌指结构(C2HC 型)(巴德仁贵等，2017)。Group 1 (D_{205}-H_{260})、Group 3 (D_{359}-H_{415})、Group 4 (D_{151}-H_{206})、Group 5 (D_{218}-H_{274} 或 P_{316}-H_{374})含有 1 个保守的 WRKY 结构域和 1 个 C_2H_2 锌指型结构域，被归为 WRKY 第 II 家族(图 7)。

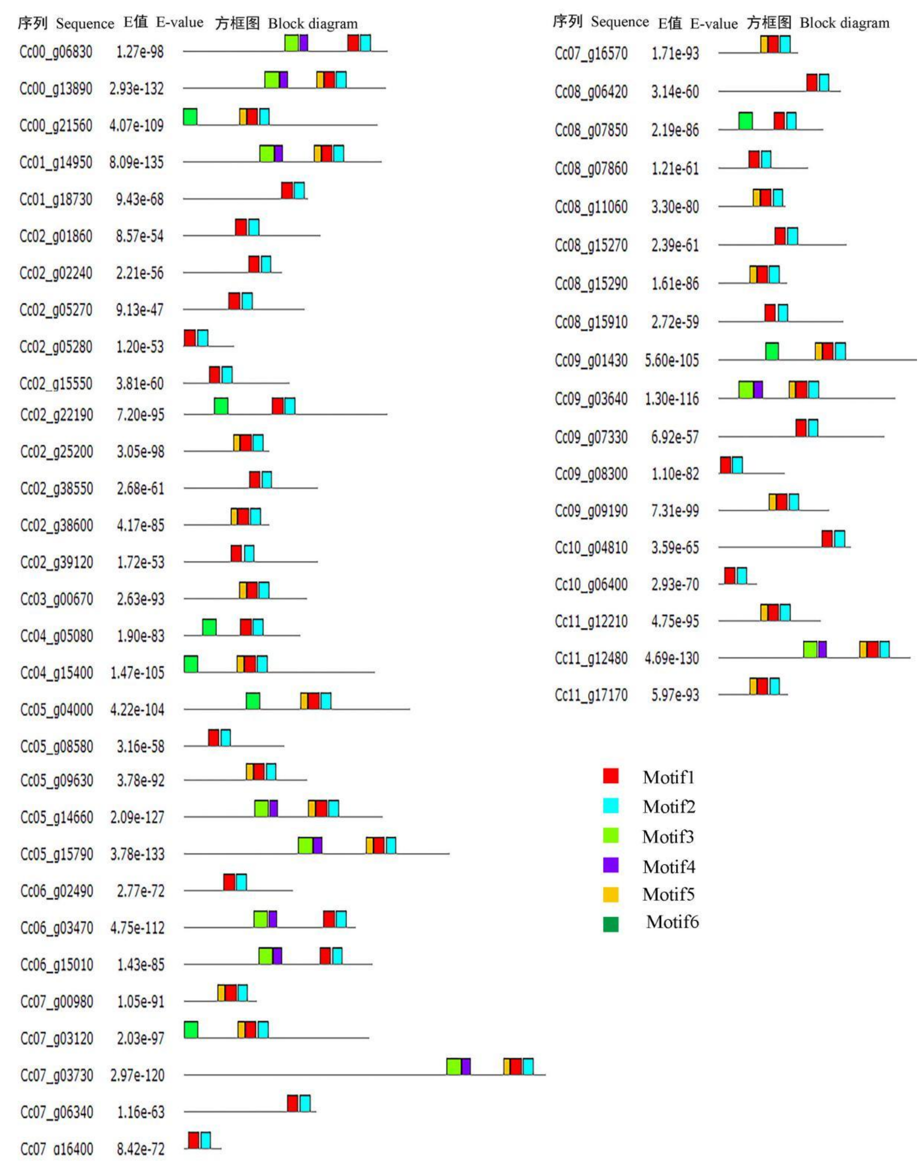


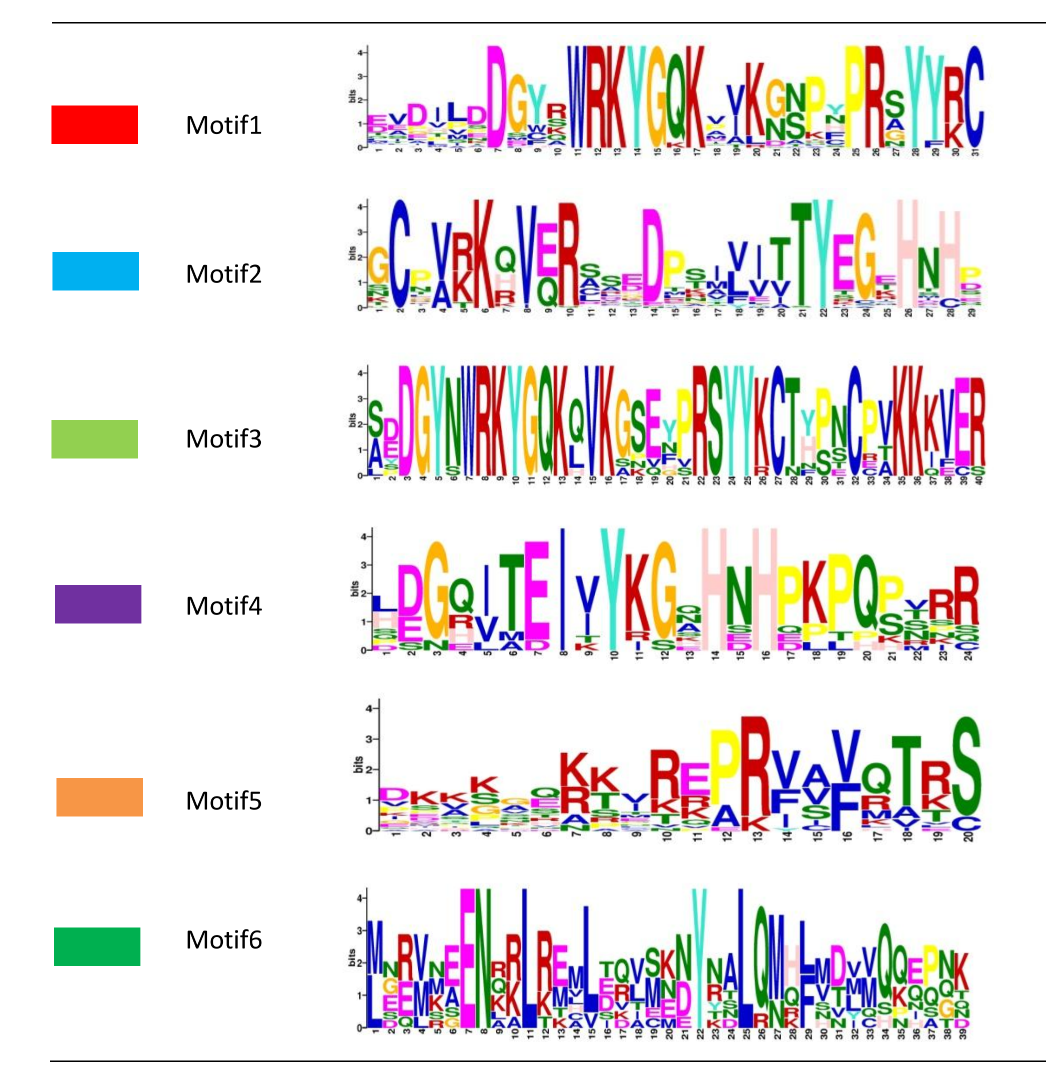
图 5 中粒咖啡 WRKY 蛋白家族的保守基序分析

Fig. 5 Conservative motif analysis of WRKY protein family in *Coffea Canephoea*

表 5 保守基序序列 logo

Table 5 Sequence logo of the conservative motifs

Motif 命名 Motif name	序列标志 Sequence logo



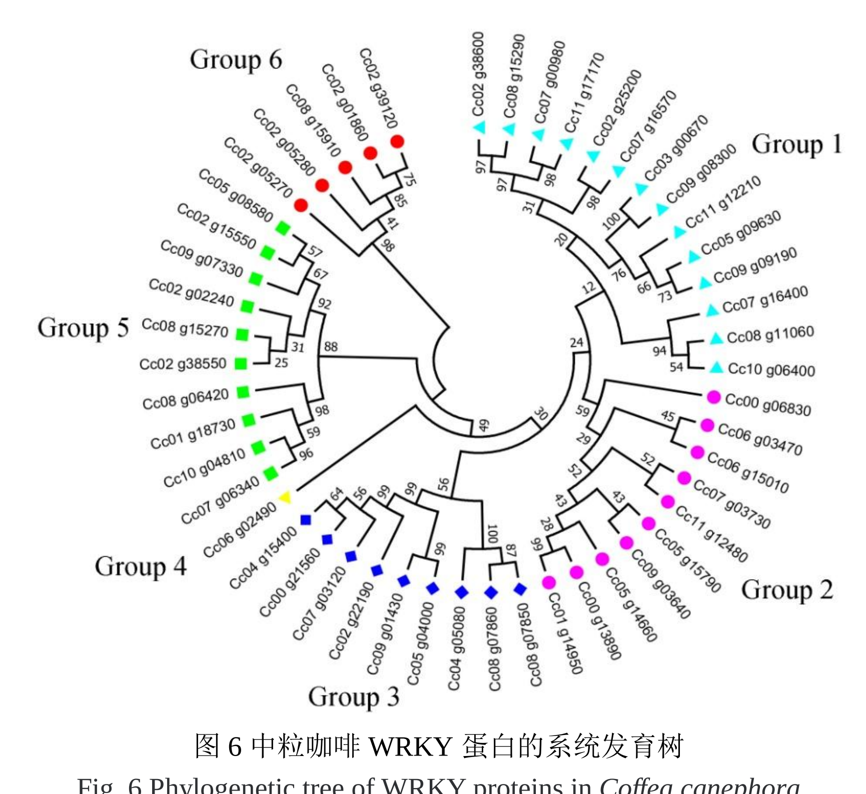


图 6 中粒咖啡 WRKY 蛋白的系统发育树

Fig. 6 Phylogenetic tree of WRKY proteins in *Coffea canephora*

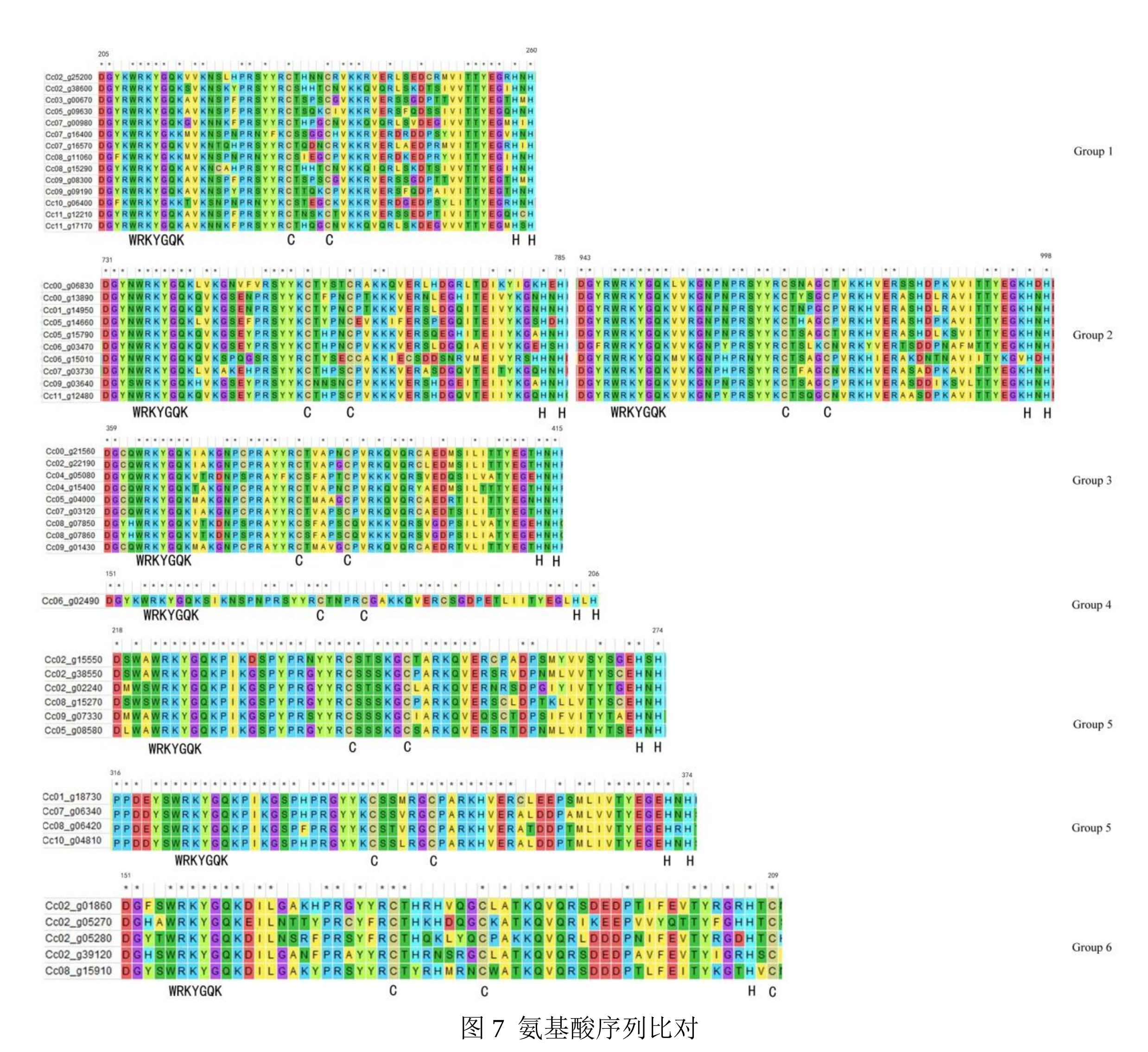


图 7 氨基酸序列比对

Fig.7 Amino acid sequence alignment

2.9 中粒咖啡转录因子 *WRKY* 与其他物种间的系统进化分析

为进一步分析中粒咖啡 *WRKY* 基因的亲缘关系和进化模式，选取拟南芥、玉米和猕猴桃等 12 个物种的转录因子 WRKY1 蛋白序列进行同源聚类分析。使用 MEGA 软件进行系统发生树的构建，进化树结果表明，中粒咖啡与 11 个物种的 *WRKY1* 转录因子家族成员可分为 2 个分支，烟草、中粒咖啡、猕猴桃、拟南芥、板栗、枣和苹果在第 1 个分支，而玉米、水稻、毛竹、小果野蕉、非洲油棕在第 2 个分支。在第 1 个分支内，中粒咖啡 *WRKY1* 与烟草亲缘关系最近，序列同源性最高；其次与猕猴桃亲缘关系较近；与非洲油棕亲缘关系最远。从双子叶角度来看，中粒咖啡与猕猴桃、烟草、拟南芥、板栗、枣和苹果这些植物聚类在一大类，符合双子叶分类标准；而单子叶玉米、水稻、毛竹、小果野蕉和非洲油棕聚类在另一大类。这表明双子叶和单子叶植物遗传进化具有很明显的分界线。从乔本类植物的角度来看，中粒咖啡 *WRKY1* 与猕猴桃的亲缘关系最近，与非洲油棕关系最远；从草本类植物来看，中粒咖啡 *WRKY1* 与烟草的亲缘关系最近，与水稻关系最远(图 8)。

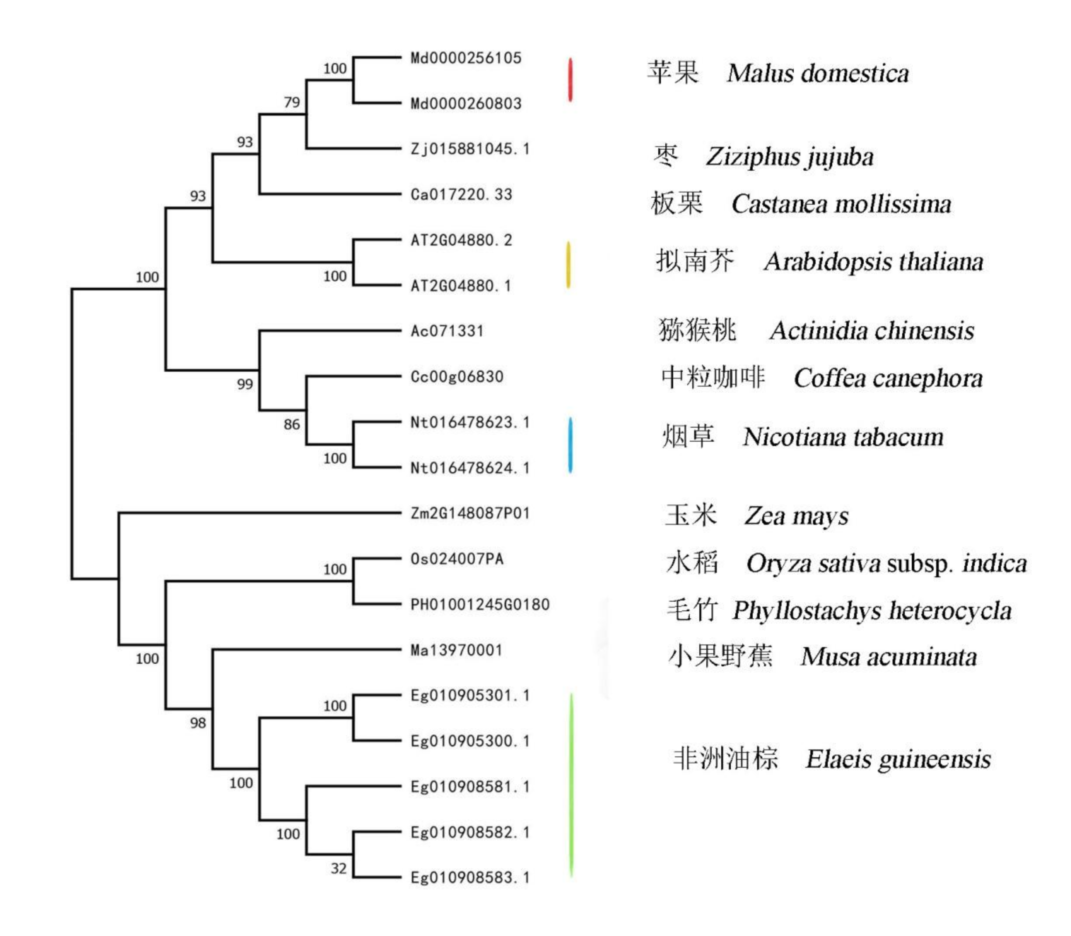


图 8 中粒咖啡与其他物种的 WRKY1 蛋白进化树分析

Fig. 8 Evolutionary tree analysis of WRKY1 protein of *Coffea canephora* and other species

3 讨论与结论

WRKY 转录因子在植物界中是通过其保守的氨基酸序列(WRKYGQK)特异识别靶基因的启动子区域 W-box [(T)TGACC(AT)]，并且调控靶基因的转录表达(梁滨和董冬，2018)。有研究报道了已经成功的鉴定和挖掘了 *AtWRKY* 和 *OsWRKY* 基因家族的生物学特征和分子功能(Rushton et al., 2010; 谷彦冰等，2015)。*WRKY* 基因家族在种子萌发的过程、植物防卫应答反应、代谢调控和发育衰老调控等生命过程中都起到重要作用(梁滨和董冬，2018)。在拟南芥 *WRKY* 抗病转录因子研究中，发现 *WRKY* 家族因子能在一定程度上提高耐受低磷胁迫的能力，对腐生病原菌起正调控作用的为 *AtWRKY3* 和 *AtWRKY4*(Lai et al., 2008; 陈益芳等，2010)；在烟草'云烟 87'叶片衰老中 *WRKY* 转录因子研究中得出对衰老的影响；烟草在植物应答病毒侵染过程研究中，推断 *NtWRKY40* 可能是参与机械损伤胁迫和病原菌侵染的重要组分(刘晶晶等，2016)。在猕猴桃涝害胁迫研究中指出 *WRKY* 在猕猴桃对涝害胁迫反应有起到重要作用(张计育等，2015)。在研究枣 *WRKY* 转录因子对枣疯病植原体和激素处理的应答中，指出枣疯病植原体侵染枣树后，诱导了 *ZjWRKY8*、*ZjWRKY52*、*ZjWRKY61* 和 *ZjWRKY69* 基因的表达(付冰等，2018)。在苹果的盐胁迫研究中得出，*MdWRKY18* 和 *MdWRKY40* 受盐胁迫的诱导，可以形成同源或异源二聚体，并增强对盐胁迫的耐性(许海峰等，2018)。总之，*WRKY* 转录因子在直接参与植物的生命活动中具有不可忽视的关键作用(Gao et al., 2016)。

本研究分析发现，中粒咖啡 *WRKY* 家族蛋白的氨基酸总数在 103~994 个之间，氨基酸平均为 376 个，等电点范围为 5.03~9.62。这与同源性较高的枣(薛宝平等，2018)、苹果(Bhattarai, 2010)*WRKY* 转录因子研究的结果是相似的，并且发现中粒咖啡 WRKY 家族蛋白都是不含跨膜结构域的亲水性稳定蛋白，都无信号肽，亚细胞定位都在细胞核中，说明该蛋白是在细胞内合成后运输到细胞核后起作用；WRKY 蛋白在丝氨酸、苏氨酸及酪氨酸处都有磷酸化位点，推测该蛋白

是通过磷酸化来实现其功能的调控。二级结构预测分析和三级结构预测分析，49 种 *WRKY* 转录因子的二级结构均主要以无规则卷曲为主，三级结构可以分为 6 类，其中以 D 类为主。有研究表明，*AtWRKY 61* 有可能参与植物免疫信号通路(Robatzek et al., 2002)；通过与 *AtWRKY72* 共同参与植物的抗病信号通路(Bhattarai，2010)；拟南芥中 *WRKY6* 参与到植物防卫反应及叶片的衰老过程中(Robatzek et al., 2002)；*AtWRKY7* 参与了拟南芥的天然免疫反应(Journot-Catalino et al., 2006)。从中粒咖啡 WRKY 蛋白的三级结构来看，D 类比较稳定，通过对蛋白序列比对拟南芥 WRKY 序列找出中粒咖啡转录因子 *WRKY* 基因家族成员的对应关系(表 1)，发现 D 类中有与 *AtWRKY61*、*AtWRKY72*、*AtWRKY6* 和 *AtWRKY7* 对应关系，推测中粒咖啡 *WRKY* 可能参与植物免疫信号通路、植物的抗病信号通路；也可能参与到植物防卫反应及叶片的衰老与天然免疫反应的过程中。由保守结构域预测分析可以看出，Cc00_g01860、Cc00_g05270、Cc00_g05280、Cc00_g39120 和 Cc08_g15910，5 个成员保守基序只含有 2 个，分别为 Motif1 与 Motif2，可以将它们分为 1 个大类，这与 WRKY 蛋白构建进化树的结果是一致的，这些成员都在 Group 6。WRKY 结构域既含保守的七肽结构(WRKYGQK)，还包含 1 个锌指结构(C2HC)，而且 C2HC 结构只存在III家族的 WRKY 结构域(李晓颖等，2019)，这与在枣等(薛宝平等，2018)其他物种上的研究结果相似。从进化分析研究表明，中粒咖啡 *WRKY* 转录因子家族蛋白序列可以分为 6 组，中粒咖啡 *WRKY1* 基因与烟草、猕猴桃和拟南芥等 11 种物种构建的系统进化树可以得到中粒咖啡 *WRKY1* 与烟草亲缘关系最近，序列同源性也较高，与非洲油棕亲缘关系最远。

本研究利用生物信息学手段在中粒咖啡基因组中发现 49 个 *WRKY* 基因，对该家族基因编码蛋白的理化性质、磷酸化位点、跨膜结构、二级和三级结构、功能结构域和分子进化等进行了详细地生物信息学分析，为后续研究中粒咖啡 *WRKY* 分子功能、进化起源关系及其逆境胁迫机制提供一定的理论依据和参考价值(巴德仁贵等，2017)。

参考文献：

BAKSHI M, OELMÜLLER R, 2014. *WRKY* transcription factors: jack of many trades in plants[J]. Plant Signal Behav, 9(2): e27700.

BAO CY, DENG L, ZHOU J, et al., 2018. Genome-wide identification and analysis of *WRKY* transcription factor family in kiwifruit[J]. Mol Plant Breed, 16:(14): 4473-4488. [包昌艳，邓浪，周军，等，2018. 猕猴桃 *WRKY* 转录因子家族全基因组鉴定与分析[J]. 分子植物育种，16(14): 4473-4488.]

BHATTARAI KK, ATAMIAN HS, KALOSHIAN I, et al., 2010. *WRKY72*-type transcription factors contribute to basal immunity in tomato and *Arabidopsis* as well as gene-for-gene resistance mediated by the tomator gene *mi-1*[J]. Plant J, 63(2): 229-240.

BUDE RG, Bao ML, ZHANG HX, et al., 2017. Bioinformatics analysis of *WRKY* transcription factor genes family in melon[J]. Genom Appl Biol, 36(11): 4761-4769. [巴德仁贵，鲍牧兰，张红霞，等，2017. 甜瓜 *WRKY* 转录因子家族基因生物信息学分析[J]. 基因组学与应用生物学，36(11)：4761-4769.]

CHEN YF, XU Q, WANG H, et al., 2010. Functional analysis of *Arabidopsis WRKY* transcription factor in plant response to low phosphorus stress[G]. Abstract Compilation of Papers on Plant Cell Biology 2010 Annual meeting of the Chinese Botany Society. [陈益芳，徐谦，王慧，等，2010. 拟南芥 *WRKY* 转录因子在植物响应低磷胁迫中的功能分析[G]. 中国植物学会植物细胞生物学 2010 年学术年会论文摘要汇编.]

CHENG T, WEI Q, LI GL, 2016. Bioinformatics analysis of the *TPS* gene family in *Coffee canephora*[J]. Chin Bull Bot, 51(2): 235-250. [程甜，魏强，李广林，2016. 中粒咖啡萜类合成酶基因家族的生物信息学分析[J]. 植物学报，51(2)：235-250.]

DAVIS AP, TOSH J, RUCH N, et al., 2011. Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and

evolutionary history of *Coffea*[J]. Bot J Lin Soc, 167(4): 357-377.
DING M, CHEN J, JIANG Y, et al., 2015. Genome-wide investigation and transcriptome analysis of the *WRKY* gene family in *Gossypium*[J]. Mol Genet Genomics, 290(1): 151-171.
DONG X, YANG Y, ZHANG Z, et al., 2019. Genome-wide identification of *WRKY* genes and their response to cold stress in *Coffea canephora*[J]. Forests, 10(4): 335.
EULGEM T, RUSHTON PJ, ROBATZEK S, et al., 2000.The *WRKY* superfamily of plant transcription factors[J]. Trends Plant Sci, 5:200-206
FU B, YE X, WANG HY, et al., 2018. Identification of *WRKY* transcription factors in jujube and their responses to '*Candidatus Phytoplasma* ziziphi' and salicylic acid or methyl jasmonate treatments[J]. Sci Silva Sin, 54(8): 65-78. [付冰，叶霞，王会鱼，等，2018. 枣 *WRKY* 转录因子的鉴定及其对枣疯病植原体和激素处理的应答[J]. 林业科学，54(8)：65-78.]
GAO R, LIU P, YONG YH, et al., 2016. Genome-wide transcriptomic analysis reveals correlation between higher *WRKY61* expression and reduced symptom severity in turnip crinkle virus infected *Arabidopsis thaliana*[J]. Sci Rep, 6(1): 24604.
GU YB, JI ZR, CHI FM, et al., 2015. Bioinformatics and expression analysis of the *WRKY* gene family in apple[J]. Sci Agric Sin, 48(16): 3221-3238. [谷彦冰，冀志蕊，迟福梅，等，2015. 苹果 *WRKY* 基因家族生物信息学及表达分析[J]. 中国农业科学，48(16)：3221-3238.]
ISHIGURO S, NAKAMURA K, 1994. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5'upstream regions of genes coding for sporamin and β-amylase from sweet potato[J]. Mol Gen Genet, 244(6): 563-571.
JIANG Y, DEYHOLOS MK, 2006. Comprehensive transcriptional profiling of NaCl-stressed Arabidopsis roots reveals novel classes of responsive genes[J]. BMC Plant Biol, 6(1): 25.
JOURNOT-CATALINO N, SOMSSICH IE, ROBY D, et al., 2006. The transcription factors *WRKY11* and *WRKY17* act as negative regulators of basal resistance in *Arabidopsis thaliana*[J]. Plant Cell, 18(11): 3289-3302.
KELLEY LA, MEZULIS S, YATES CM, et al., 2015. The Phyre2 web portal for protein modeling, prediction and analysis[J]. Nat Protoc, 10(6): 845-858.
KUMAR S, STECHER G, LI M, et al., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms[J]. Mol Bio Evol, 35(6): 1547-1549.
LAGACÉ M, MATTON DP. 2004. Characterization of a *WRKY* transcription factor expressed in late torpedo-stage embryos of *Solanum chacoense*[J]. Planta, 219(1): 185-189.
LAI Z, VINOD KM, ZHENG Z, et al., 2008. Roles of arabidopsis *WRKY3* and *WRKY4* transcription factors in plant responses to pathogens[J]. BMC Plant Biol, 8(1): 68.
LI XY, XU HX, CHEN JW, 2019. Identification and expression analysis of *WRKY* transcription factors in *Eriobotrya japonica*[J]. Acta Hort Sin, 46(5): 939-954. [李晓颖，徐红霞，陈俊伟，2019. 枇杷 *WRKY* 转录因子鉴定与表达分析[J]. 园艺学报，46(5)：939-954.]
LI Z, HAN X, ZHANG YH, et al., 2019. Cloning and functional analysis of a phosphorus stress response gene *GhWRKY6*, in cotton (*Gossypium hirsutum* L.) [J]. Cotton Sci, 31(3): 182-191. [李智，韩笑，张一豪，2019. 棉花磷胁迫响应基因 *GhWRKY6* 克隆及功能分析[J]. 棉花学报，31(3)：182-191.]
LIANG B, DONG D，2018. Advances in *WRKY* transcription factors in plants[J]. Bull Biol, 53(6): 5-8. [梁滨，董冬，2018. 植物 *WRKY* 转录因子的研究进展[J]. 生物学通报，53(6)：5-8.]
LIU JJ, CHENG CL, XI YZ, et al., 2016. Roles of *Nicotiana tobacum NtWRKY40* in plant responding to virus infection[J]. Biotechnol Bull, 32(10): 188-198. [刘晶晶，程春玲，席玉珍，等，2016. 烟草 *NtWRKY40* 在植物应答病毒侵染过程中的作用[J]. 生物技术通报，32(10): 188-198.]
MONENTE C, LUDWIG IA, IRIGOYEN A, et al., 2015 Assessment of total (free and bound) phenolic

compounds in spent coffee extracts[J]. J Agric Food Chem, 63(17): 4327-4334.
ROBATZEK S, SOMSSICH IE, 2010. A new member of the *Arabidopsis WRKY* transcription factor family, *AtWRKY6*, is associated with both senescence- and defence-related processes[J]. Plant J, 28(2): 123-133.
ROBATZEK S, SOMSSICH IE, 2002. Targets of *AtWRKY6* regulation during plant senescence and pathogen defense[J]. Gene Dev, 16(9): 1139-1149.
RODRIGUES NP, SALVA TDJ, BRAGAGNOLO N, 2015. Influence of coffee genotype on bioactive compounds and the in vitro capacity to scavenge reactive oxygen and nitrogen species[J]. J Agric Food Chem, 63(19): 4815-4826.
RUSHTON DL, TRIPATHI P, RABARA RC, et al., 2012. *WRKY* transcription factors: key components in abscisic acid signalling[J]. Plant Biotechnol J, 10(1): 2-11.
RUSHTON PJ, SOMSSICH IE, RINGLER P, et al., 2010. *WRKY* transcription factors[J]. Trends Plant Sci, 15(5): 247-258.
SAITOU N, NEI M. 1987.The neighbor-joining method: a new method for reconstructing phylogenetic trees[J]. Mol Biol Evol, 4(4): 406-425.
SHA T, LI YJ, ZHANG FB, et al., 2019. Antigenic epitopes of echinococcus granulosus proteins: egG1y162-1 and egG1y162-2: a comparative study[J]. J Xinjiang Med Univ, 42(1): 12-19.[沙桐，李玉娇，张峰波，等，2019. 细粒棘球绦虫蛋白 egG1y162-1, egG1y162-2 的抗原表位比较[J]. 新疆医科大学学报，42(1)：12-19.]
SU L, WANG PF, YANG Y, et al., 2019, Genome-wide identification and analysis of *WRKY* transcription factors in grape[J]. Heilongjiang Agric Sci, (1): 13-22.[苏玲，王鹏飞，杨阳，等，2019. 葡萄全基因组 *WRKY* 转录因子鉴定和分析[J] 黑龙江农业科学，(1)：13-22.]
WANG PY, QU SS, LIANG Y, et al., 2019. Bioinformatics analysis of *WRKY* transcription factor family in rice[J]. J Anhui Agr Sci, 47(12):123-126.[王鹏洋，曲姗姗，梁源，等，2019. 水稻 *WRKY* 转录因子家族的生物信息学分析[J]. 安徽农业科学，47(12)：123-126.]
WU X, SHIROTO Y, KISHITANI S, et al., 2009. Enhanced heat and drought tolerance in transgenic rice seedlings overexpressing *Oswrky11* under the control of *hsp101* promoter[J]. Plant Cell Rep, 28(1): 21-30.
XIE ZW, WANG LJ, CHEN JY, et al., 2016. Studies on *WRKY* transcription factors and their biological functions in plants[J]. J Agric Sci Technol China, 18(3): 46-54.[谢政文，王连军，陈锦洋，等，2016. 植物 WRKY 转录因子及其生物学功能研究进展[J]. 中国农业科技导报，18(3)：46-54.]
XU HF, YANG GX, ZHANG J, et al., 2018. Molecular mechanism of apple *MdWRKY18* and *MdWRKY40* participating in salt stress[J]. Sci Agric Sin, 51(23): 4514-4521.[许海峰，杨官显，张静，等，2018. 苹果 *MdWRKY18* 和 *MdWRKY40* 参与盐胁迫途径分子机理研究[J]. 中国农业科学，51(23)：4514-4521.]
XU YH, WANG JW, WANG S, et al., 2004. Characterization of *GaWRKY1*, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-δ-cadinene synthase-A1[J]. Plant Physiol, 135(1): 507-515.
XUE BP, ZHANG H, ZU HH, et al., 2018. Bioinformatics analysis of the *WRKY* transcription factor family in *Ziziphus jujuba*[J]. Mol Plant Breed, 16(14): 4543-4558. [薛宝平，张昊，祖欢欢，等，2018. 枣 *WRKY* 转录因子家族的生物信息学分析[J]. 分子植物育种，16(14)：4543-4558.]
YAN L, HUANG LF, WANG XY, et al., 2018. Comparative trial of robusta coffee cultivars[J]. Chin J Trop Crop, 39(7): 1276-1281.[闫林，黄丽芳，王晓阳，等，2018. 中粒种咖啡品种比较试验[J]. 热带作物学报，39(7)：1276-1281.]
YANG HG, 2014. Effects of drought stress on lipid peroxidation, antioxidant enzyme activities and osmolyte content in leaves of *Coffea canephora* seedlings[J]. Chin J Trop Crop, 35(5): 944-949.[杨华

庚，2014. 干旱胁迫对中粒种咖啡幼苗膜脂过氧化、抗氧化酶活性和渗透调节物质含量的影响[J]. 热带作物学报，35(5)：944-949.]

YU CS, CHEN YC, LU CH, et al., 2006. Prediction of protein subcellular localization[J]. Proteins, 64(3): 643–651.

ZHANG CL, MAN LL, XIANG DJ, et al., 2018. Cloning and bioinformatics analysis of *BjSnRK2C* gene from *Potherb mustard*[J]. Acta Agric Boreal-Sin, 33(6): 24-32. [张春兰，满丽莉，向殿军，等，2018. 雪菜 *BjSnRK2C* 基因的克隆及生物信息学分析[J]. 华北农学报，33(6)：24-32.]

ZHANG H, 2017. Bioinformatics analysis of *WRKY* transcription factor family and identification of some genes related to stress resistance in tomato[D]. Harbin: Northeast Agricultural University.[张红，2017. 番茄 *WRKY* 转录家族生物信息学分析及部分抗逆相关基因鉴定[D]. 哈尔滨：东北农业大学.]

ZHANG JY, HUANG SN, MO ZH, et al., 2015.Transcriptional sequencing and comparative analysis of the different expression genes of *Actinidia chinensis* under the drought stress[C]. Annual meeting of Chinese Society for Horticultural Science in 2015.[张计育，黄胜男，莫正海，等，2015. 涝害胁迫下猕猴桃差异表达基因的转录组测序及比较分析[C]. 中国园艺学会 2015 年学术年会.]

ZHANG L, HE XJ, LI C, et al., 2019. Expression and functional prediction of *WRKY* transcription factors in leaf senescence of tobacco ‘Yunyan 87’[J]. Mol Plant Breed, 17(8).[张柳，何晓健，林春，等，2019. 烟草‘云烟 87’叶片衰老中 *WRKY* 转录因子的表达及功能预测[J]. 分子植物育种，17(8).]

ZHENG C, ZHENG ES, WANG YM, et al., 2018. Research progress on rice *WRKY* transcription factors[J]. Letters in Biotechnol, 29(2): 286-294.[郑超，郑二松，王栩鸣，等，2018. 水稻 *WRKY* 转录调控因子研究进展[J]. 生物技术通讯，29(2)：286-294.]

ZHOU J, HUANG J，2018. Cloning and functional identification of chitinase gene *SlChi* in Salix[J]. Mol Plant Breed, 16(24): 8013-8021.[周洁，黄婧，2018. 柳树几丁质酶基因 *SlChi* 的克隆和功能验证[J]. 分子植物育种, 16(24)：8013-8021.]

ZOU Z, 2013. Genome-wide identification and phylogenetic analysis of *WRKY* transcription factor family in castor bean[J]. Oil Crop Sci, 35(1): 36-42.[邹智，2013. 蓖麻 *WRKY* 转录因子的全基因组鉴定及其进化分析[J]. 中国油料作物学报, 35(1)：36-42.]